Nitrotoga is selected over Nitrospira in newly assembled biofilm communities from a tap water source community at increased nitrite loading

Kinnunen, Marta; Gülay, Arda; Albrechtsen, Hans-Jørgen; Dechesne, Arnaud; Smets, Barth F.

Published in:
Environmental Microbiology

Link to article, DOI:
10.1111/1462-2920.13792

Publication date:
2017

Document Version
Peer reviewed version

Title:

Nitrotoga is selected over *Nitrospira* in newly assembled biofilm communities from a tap water source community at increased nitrite loading

Short title:

Ecological processes in NOB guild assembly

Author affiliations:

Marta Kinnunen, Arda Gülay, Hans-Jørgen Albrechtsen, Arnaud Dechesne, Barth F. Smets*

Technical University of Denmark, Department of Environmental Engineering, Bygningstorvet 115, 2800 Kgs. Lyngby, Denmark

*correspondence to: Barth F. Smets; Technical University of Denmark; DTU Environment; Bygningstorvet 115; 2800 Kgs. Lyngby; Denmark; Ph. +45 45 25 22 30; Fax +45 45 93 28 50; E-mail bfsm@env.dtu.dk

Originality-Significance Statement.

We characterize experimentally community assembly processes in newly assembled biofilm communities. We show how both total community and guild-level analyses provide evidence for contribution of neutral processes combined with selection. We investigate, for the first time, how environmental filtering caused by elevated nitrite loading results in the dominance of *Nitrotoga* over *Nitrospira* in nitrite-oxidizer guild.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13792

Summary

Community assembly is a central topic in microbial ecology: how do assembly processes interact and what is the relative contribution of stochasticity and determinism? Here, we exposed replicate flow-through biofilm systems, fed with nitrite-supplemented tap water, to continuous immigration from a source community, present in the tap water, to determine the extent of selection and neutral processes in newly assembled biofilm communities at both the community and the functional guild (of nitrite-oxidizing bacteria, NOB) levels. The community composition of biofilms assembled under low and high nitrite loading was described after 40 days of complete nitrite removal. The total community assembly, as well as the NOB guild assembly were largely governed by a combination of deterministic and stochastic processes. Furthermore, we observed deterministic enrichment of certain types of NOB in the biofilms. Specifically, elevated nitrite loading selected for a single *Nitrotoga* representative, while lower nitrite conditions selected for a number of *Nitrospira*. Therefore, even when focusing on ecologically coherent ensembles, assembly is the result of complex stochastic and deterministic processes that can only be interrogated by observing multiple assemblies under controlled conditions.

Introduction

Managing and steering complex microbiological systems requires an understanding of microbial community assembly. In order to conceptually organize seemingly disparate theories and concepts in community ecology, Vellend (2010) suggested focusing on the four processes that can be considered the basic elements of any community assembly theory: selection, drift, dispersal and speciation. The relative importance of these processes is under debate and may vary with community successional stage, size, and type. Community assembly processes are often also classified as deterministic or stochastic: deterministic processes such as selection are mostly driven by environmental conditions or species interactions, whereas stochastic (or neutral) processes are driven by chance (e.g. drift) and

unpredictable disturbances (Stegen *et al.*, 2013). Some processes have both deterministic and stochastic components (e.g. dispersal) (Evans *et al.*, 2017). Other frameworks distinguish between local (environmental conditions) and regional (historical events, dispersal from a meta-community) processes as shaping microbial community assembly (Lindström and Langenheder, 2012).

Theoretical frameworks ultimately need to be confronted with experimental observations and several studies have tried to quantify the relative importance of different processes (e.g., deterministic *vs* neutral or local *vs* regional) in shaping community assembly: these studies typically compare established microbial communities across various spatial or ecological scales (microbial biogeography, β–diversity studies (Wang *et al.*, 2013)) or examine *de novo* assembly of replicate microbial communities in batch microcosms (wherein the source meta-community and/or local conditions are varied) (Lee *et al.*, 2013; Moroenyane *et al.*, 2016). Occasionally, the dominance of one process over another is evident: Moroenyane *et al.* (2016) shows dominance of local factors while Burns *et al.*, (2015) shows dominance of stochastic factors; more typically multipe processes are shown to interact or cannot be uncoupled (e.g. Ofiteru *et al.* (2010) and Langenheder and Székely (2011)).

We argue that most current empirical approaches are unnecessarily complex: deterministic and stochastic processes can seldom be untangled in studies of established communities (Moroenyane *et al.*, 2016); on the other hand, studies comparing newly assembled communities are performed in batch microcosms where conditions are time variant and community assembly processes are unlikely at equilibrium (Lee *et al.*, 2013). Finally, while the neutral community assembly model (NCM) provides an accessible tool to measure the significance of stochastic processes in community assembly (Sloan *et al.*, 2006), most experimental systems violate the central NCM assumption of functional equivalence among

all community members (Burns *et al.*, 2015). While only experiments using one generalist guild have been reported (De Roy *et al.*, 2013), communities are in general complex with several guilds (for example Langenheder and Szekely, 2011).

Our aim was to evaluate the role of stochastic *vs* deterministic processes in community assembly. Since it is difficult to ascertain selection - and apply the NCM - in complex communities, we focused on guild assembly: members of a guild are, by definition, functionally similar and guilds can be easily established by supplementation with a specific energy source during assembly. Hence, in a system where the potential for stochasticity is high (initial assembly – Langenheder and Székely (2011)) we enforce a constant enrichment pressure for a specific guild to determine the relative effects of stochastic and deterministic processes.

As model guild, we employed nitrite-oxidizing bacteria (NOB). The diversity of NOB is remarkable, with genera distributed across multiple bacterial phyla (Nitrospinae, Nitrospirae, Proteobacteria, Chloroflexi) and lineages (*Nitrobacter, Nitrococcus, Nitrospina, Nitrospira, Nitrotoga, Nitrolancetus)* (Ward *et al.*, 2011; Sorokin *et al.*, 2012). The maintenance of this diversity suggests that NOB have different niches. Nitrite concentration has been suggested as the main determinant of niche difference (Nowka et al., 2015; Daims et al., 2016), even though the evidence derives primarily from pure culture observations and not from *in situ* guild inspections under different nitrite conditions (but see Maixner *et al.*, (2006)). Temperature preferences (e.g. for psychrophilic *Nitrotoga* – Alawi *et al.* (2009)) and additional metabolic capabilities (e.g. ability to use ammonium (Daims *et al.*, 2015, 2016; Pinto *et al.*, 2015; Palomo *et al.*, 2016), hydrogen (Koch *et al.*, 2014), and urea (Koch *et al.*, 2015) as substrate by various *Nitrospira* lineages) have also been implicated in niche differentiation among NOBs. This provides a base for environmental filtering (i.e. selection) to act in NOB guild assembly.

To evaluate the strength of deterministic (selection) processes compared to stochastic (drift, dispersal) processes at play in NOB guild assembly, we assembled, under two defined nitrite loadings, replicate biofilms from a meta-community immigrating from tap water at constant rate. We hypothesized that the NOB guild composition of biofilm communities assembled under the two loading rates would consistently differ from each other and from the source meta-community (indicating deterministic processes) and we expected that assembly of the non-NOB taxa (entire community) would show more stochasticity than the NOB guild. In addition to identifying the dominance of selection, we also aimed at describing the NOB types enriched under the two loadings, and hypothesized that *Nitrospira* types would dominate the biofilms at low loading regimes, and *Nitrobacter* types would dominate at high loading regimes.

Results

Nitritatation performance

Nitrite removal commenced rapidly in both systems, and was complete from day 22 onwards; nitrite was stoichiometrically converted to nitrate (Figure S1). Combined, these results indicate facile enrichment of NOBs. At 63 days, after 40 days of performance at steady state, biofilms were sacrificed.

Community density

Biofilms formed at high nitrite loading rate contained in average (± standard error of the mean) $2.7x10^6 \pm 2.3x10^6$ cells per reactor while lower loading conditions resulted in significantly less dense biofilms ($p<0.001$) with an average of $6.78x10^5 \pm 8.95x10^5$ cells per reactor according to qPCR targeting the 16S rRNA gene (Table S1). The average immigration from the source community, derived from the average density of cells in the tap water, was $1.90x10^8 \pm 2.20x10^8$ cells/day (Table S2).

Biofilm community-level composition

Comparative analysis of the rarefied 16S rRNA gene amplicons libraries showed significant shifts of richness, diversity and evenness from the source to the biofilm communities (Figure 1A and Table S3). The observed and estimated richness was highest in the source community, mainly because of the presence of rare taxa (<1% in relative abundance). However, the richness was not significantly different between the two biofilm communities. Consistently, the source community was more even (see Gini coefficient, Figure 1A) than the assembled biofilm communities, where a few taxa became highly abundant. When the biofilms were compared to each other, the lowest α-diversity, as well as highest unevenness, were estimated for the high nitrite loading (HNL) communities.

It was also evident that the A composition shifted from the source community to the two biofilm communities (Figure 2). The most abundant community members in the source (*Burkholderiales* and α-proteobacteria) decreased in relative abundance in the biofilm community assembly. The largest increases in biofilms were sequence variants corresponding to different NOB, mostly at the expense of unclassified α-proteobacteria that made up about 10% of the source community, and of unclassified bacteria that made up about 20%. In addition, the low nitrite loading (LNL) condition caused a consistent increase in the relative abundance of *Rhodospirillales*, which was not observed under HNL conditions. Similarly, members of the *Chromatiales* became especially abundant in the HNL biofilms compared to the source and the LNL communities.

NOB guild composition

All sequence variants corresponding to NOB were classified as *Nitrospira* or *Nitrotoga* (Figure 2 and Figure 3). No *Nitrobacter* nor *Nitrococcus* (other typical NOB genera) sequences were detected in any of the samples. Sequence variants classified as NOB made up less than 10% of the source community and were enriched to nearly half of the biofilm communities.

Even though we observe difference in richness at the NOB guild level in the assembled biofilms (Figure 1B, Chao 1), that difference is not statistically significant (Table S3). Significant differences at the guild level were observed for evenness ($p<0.0001$) (Figure 1B Gini index and Table S3) and diversity (estimated by Shannon index) ($p<0.005$) (Figure 1B and Table S3). Following the same pattern as for total community diversity, the NOB guild diversity was higher in the LNL than the HNL biofilms ($p=0.0008$).

In the source community, *Nitrotoga* and *Nitrospira* were both present at low abundances (Figure 3) while in the LNL biofilm communities, *Nitrospira* were the most dominant NOB. The most abundant *Nitrospira* sequence variants belonged to Lineage 2 (Figure S3), with high relatedness to *Nitrospira moscovensis*, *Ca.* Nitrospira nitrosa isolate and *Nitrospira lenta*. In the HNL biofilm communities, on the other hand, *Nitrotoga* became the most abundant NOB in most replicates, among which two *Nitrotoga* sequence variants were dominant (Figure S3). The dominant *Nitrotoga* sequence variant shared high level of identity with uncultured *Ca*. Nitrotoga sp. clones (99.5 %) (Figure S3).

Community assembly processes

We compared the biofilm community compositions to those expected based on the neutral community assembly model (Sloan *et al.*, 2006); first, we used only taxa that were shared in the source and the biofilm communities. This approach, commonly used in NCM, implies that the sequences that are not present in both the source and local communities are excluded from the analysis. In our case, more than half of the sequences were thus excluded. To avoid that our evaluation of the NMC is only based on a fraction of the information obtained from amplicon sequencing, we also used a second approach, where we considered all taxa present in any of the biofilms to be also present in the source, even when it was not detected (from here on referred to as extended source community). We

assigned the lowest observed average relative abundance in the source community (0.001%) to those sequence variants in the source community.

When focusing only on the shared sequence variants, we clearly rejected the hypothesis that there is a significant correlation between the model prediction and the observed data at both the total community and the NOB guild levels (Table S4, Figure S4 and Figure S5). We observed high correlation coefficient for the LNL NOB but, due to low number of sequence variants that were shared between the biofilm and the source community (7 sequence variants); this correlation was statistically not significant (Table S4). Only after extending the source community did we see significant support for neutral assembly. At the total community level, the correlation between the sequence variant frequencies observed in the biofilms and their mean relative abundance in the extended source community ($\rho_{Spearman}$ =0.39 and 0.46, for LNL and HNL, respectively (Table S4; Figure S4)) suggested that there was a minor contribution of neutral processes in biofilm assembly. This contribution was shown to be higher at NOB guild level, where we observed correlations $\rho_{Spearman}$ =0.49 and 0.51, for LNL and HNL, respectively (Table S4; Figure S5).

In order to elaborate on the levels of stochasticity in the biofilm communities, we evaluated the dissimilarities in the composition (using Bray-Curtis distances) within and across replicate assembled biofilms, and source communities (Evans *et al.*, 2017).

The distances between biofilms and source communities indicate that deterministic processes influenced the total biofilm community assembly more than the NOB guild assembly (Figure 4). The HNL biofilm communities show large within-group distances, representing the influence of stochastic processes, while this dissimilarity is lower in LNL biofilm communities. The within-treatment dissimilarities at the NOB guild level are similar for HNL and LNL.

Discussion

Disentangling the importance of different processes is the goal of most community assembly studies. It is crucial to recognize that particular processes may have more or less importance depending on the scenario and might therefore require complementary methods for correct detection and quantification. Here, we examined the scenario of newly assembled communities, with focus on a single functional guild, with continuous dispersal, under different defined and constant resource loadings enriching for the said guild. Previous studies on initial community assembly, using batch microcosms, revealed that the first colonization steps were critical, with a suspected role of priority effect, which has a high stochastic component (Burns *et al.*, 2015). However, when succession could proceed under dispersal from a meta-community, the importance of this largely stochastic initial assembly steps declined and the influence of selection grew larger (Langenheder and Szekely, 2011; Lee *et al.*, 2013).

The composition of the tap water community described over several weeks varied very little, indicating that this constituted a stable source community (Figure 2, 4 and S2). The assembled biofilm communities deviated significantly from this source community (Figure 2, 4, S2 and Table S3). Resource addition (nitrite-N) was primarily responsible for this compositional change, as sequence variants belonging to NOB became dominant under both HNL and LNL scenarios (Figure 2), resulting in significant reduction in community richness and evenness (Figure 1). The within-group dissimilarity of repeatedly assembled HNL and LNL communities exceeded that of the repeatedly sampled source community (Figure 4), which highlights a stochastic contribution to community assembly (Evans *et al.*, 2017). Support for neutral assembly at the total community level was also significant when the source community was extended to include all exact sequence variants detected in the biofilms, even when said exact sequence variant was un-detected in the source community. However, the contribution of determinism (selection) at the total community level was

evident in the large and consistent dissimilarities between the source and assembled communities (Figure 4).

Major changes in the total biofilm community composition were expected due to the intentional provision of nitrite, resulting in the enrichment of the NOB guild, the assembly of this guild was monitored at the two nitrite loads to evaluate the presence and strength of environmental filtering. The NOB guild composition of the biofilm communities differed significantly from that of the source community (Figure 3), as most evident in the relative abundance of *Nitrospira* and *Nitrotoga* sequence variants. *Nitrospira* and *Nitrotoga* were at low abundance in the source community, but the NOB guild was consistently skewed towards *Nitrospira* dominance (18.7%, Figure 3) and *Nitrotoga* dominance (39.7%, Figure 3) in the LNL and HNL biofilms, respectively. While the evenness and richness of the NOB guild appeared to increase for the LNL biofilms, they dropped in the HNL biofilms (Figure 1), in large part due to the dominance of two *Nitrotoga* sequence variants (Figure S3).

This differential dominance pattern indicates that the competitive advantage of *Nitrotoga* over *Nitrospira* is not restricted to low temperatures, as suggested in Alawi *et al.* (2009), but also includes higher nitrite loading rates. *Nitrospira* dominance over *Nitrotoga* at lower nitrite loading rates is consistent with *Nitrospira*'s high affinity for nitrite (Nowka *et al.*, 2015) – even though under both loading conditions, nitrite concentrations were (from day 20 onward) below detection limit (0.028 mgN/L). The dominance of Lineage 2 over other *Nitrospira* lineages in LNL is consistent with other observations (Maixner *et al.*, 2006; Nowka *et al.*, 2015), yet its replacement by *Nitrotoga* (and not other *Nitrospira* lineages or *Nitrobacter*) at elevated loading rates has not been reported before. Obviously, environmental filtering only acts on guild members present in the source community: in our case, only *Nitrotoga* and Lineage 2 *Nitrospira*. It is therefore possible that other NOB could displace *Nitrotoga* in the HNL biofilms, if present in the source.

While strong environmental filtering in the NOB guild was evident in the compositional analysis of replicate communities (Figure 4), we also noted multiple evidence of stochasticity. First, guild composition correlated with the NCM in both loading regimes (ρ =0.506 and ρ =0.487 for HNL and LNL, respectively) when considering the extended source community, similarly to total community. These significant correlation values suggest that the frequency of detection of NOB sequence variants in the biofilm largely followed that of expected, based solely on their abundance in the tap water (source community) (Figure S5), especially for sequence variants that were rare in the source community. However, support to the NCM disappeared when we focused only on the sequence variants that were shared between the source and the biofilms. This suggests that including the rare taxa in the source community (not detected with our sampling method or sequencing) increases the support of the NCM. Others report much higher support of the NCM but in all these studies composition of the source community is not directly measured but estimated by averaging the composition of multiple communities (Sloan *et al.*, 2006; Ofiteru *et al.*, 2010; Burns *et al.*, 2015; Roguet *et al.*, 2015). We believe that such approach may provide misleading confidence in the NCM. Second, the within-group dissimilarity of repeatedly assembled HNL and LNL NOB guilds exceeds the dissimilarity of the repeatedly sampled source community (Figure 4); however, this dissimilarity is smaller at guild level, than at the total community level; and, third, there is substantial overlap between the LNL and HNL guild compositions (Figure 4).

The *Nitrotoga* taxon that was selected in HNL was a very minor member of the source community. It is therefore likely that, by chance, it did not colonize early some of the biofilms and thus failed to reach dominance at the time of sampling (as observed in three replicate HNL biofilms, Figure 2). This highlights how environmental filtering also depends on stochastic events such as dispersal. Alternatively, it is possible that the selective advantage

of *Nitrotoga* over the rest of the guild members was not high enough to result in its dominance in all replicates. In any case, the low diversity of NOB in the HNL biofilms compared to the LNL biofilms, associated to the extreme dominance of a single sequence variant, suggest that very few source community members possessed a selective advantage under high nitrite loading. This is not unexpected as the source community is a drinking water distribution system community, which has likely developed exposed to low nitrogen loading.

In conclusion, our work highlights the importance of extending the community assembly investigation to specific guilds. The presence-absence of sequence variants in the biofilms, compared to the source community indicated a significant fit with the NCM both at total community and at the guild level, being higher at the NOB guild level, only when an extended meta-community approach was used. In contrast, when only shared sequence variants were considered both total community and the guild assembly showed no correlation with the NCM prediction. Guild assembly under low nitrite loading elicited, compared to the source, increased dominance of *Nitrospira*, whereas assembly under high nitrite loading selected for *Nitrotoga* sequence variants. However, the fact that the dominance of *Nitrotoga* sequence variants was not observed in all replicates, suggests that the environmental filtering was only moderate, either because of its very low abundance in the source community or because its fitness differential with other NOB was low. Therefore, the combination of compositional analysis of replicate communities and constituent guilds assembled under different environmental conditions can provide rich information on the processes governing community assembly.

Experimental procedures

Experimental set-up

The experimental set-up consisted of 40 parallel flow-through silicone tubes. Biofilms developed on the inner surface (surface area 8.84×10^{-4} m^2) of the tubes by feeding with

regular tap water spiked with nitrite at a constant flow rate of 0.43 L/day under ambient temperatures (23 to 25°C). Twenty replicates were fed tap water with 1mgN/L nitrite addition while 20 other replicates received 10-fold lower nitrogen concentration, 0.1 mgN/L. See SI for methods of nitrogen measurements. Biofilm was allowed to develop in the system for 63 days after which the biofilm from the tubes was extracted by enforcing high shear force with saline buffer to detach the biofilm witch was stored at -20°C for subsequent DNA analysis.

DNA extraction

DNA from the extracted biofilm and tap water cells (sampled by filtering 2L through 0.1µm filter once a week over a 6 week period) was isolated using the FastDNA™ SPIN Kit for Soil and the FastPrep® Instrument (MP Biomedicals, Santa Ana, CA) according to the manufacturer's instruction at room temperature. The concentration and purity of extracted DNA were checked using NanoDrop™ 2000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE). DNA was then stored at -20°C for subsequent molecular analyses.

qPCR

Real-time qPCR assays were performed with a Chromo4 Thermal cycler (Bio-Rad Laboratories, Hercules, CA). Reaction mixtures (25 µl) contained 12.5 µl SYBR® Green qPCR Mastermix (iQ™ SYBR® Green Supermix; Bio-Rad, Hercules, CA) 1 µl forward and reverse primers (20 µM), 5 µl of template DNA (adjusted to 2 ngDNA μl^{-1}) and 5.5 µl PCR-grade water. Total bacteria were quantified based on 16S rRNA gene copy numbers using the Eubacterial primer set 1055f-1392r as described in Terada *et al.*, (2010). On average 2.5 copies of 16S rRNA gene was estimated per cell, according to *rrn*DB (Stoddard *et al.*, 2015), with the assumption that majority of the community belongs to *Gallionellacea* and *Nitrospiraceae*. *Nitrospira* cells were quantified using Nitrospira-specific qPCR with primer set NTS232f (Lim et al., 2008) and Nsr1264r (Dionisi et al., 2002) targeting the 16S rRNA genes. Cell numbers were calculated assuming a single 16S operon copy per cell (*rrn*DB).

Sequencing of 16S rRNA amplicons

Extracted DNA from the biofilm and the tap water samples were PCR-amplified using primer set PRK341F (5'- CCTAYGGGRBGCASCAG-3') and PRK806R (5'-GGACTACNNGGGTATCTAAT-3') for 35 cycles, to amplify the V3-V4 hypervariable regions (Yu *et al.*, 2005). Purified PCR products were sequenced on the Illumina MiSeq platform at the DTU Multi Assay Core Center (Lyngby, DK).

All raw 16S rRNA gene amplicons were processed following the DADA2 pipeline with default settings (Callahan et al., 2016). DADA2 is an open-source R package that uses a model-based approach for correcting amplicon errors without constructing OTU-s, instead exact sequence variants are reported. These sequence variants were classified based on the SILVA prokaryotic reference database version 123. A total of 2.6 million sequences passed the filtering step, amounting to 62 000 sequences per sample on average (46 samples) (ranging from 9400 to 190 000 reads per sample). Five samples with less than 5000 sequences were removed from downstream analysis. See supplemental methods for criteria for assigning sequences to NOB guild. All sequences have been submitted to NCBI Sequence Read Archive under accession number SRP092237.

Neutral community assembly model (NCM)

To determine the importance of neutral processes in community assembly, the model described by Sloan *et al.* (2006) was applied. This neutral community assembly model (NCM) predicts the frequencies with which taxa should occur in local communities based on their abundance in the source community. In short, the model predicts that taxa that are abundant in the source community will be observed more frequently in the local communities due to increased dispersal opportunities, while rare taxa are more likely to be lost in the local community due to ecological drift.

The model fit is determined by a single free parameter – the migration rate. It is the estimated probability that random loss of an individual in local community will be replaced

by dispersal from the source community. The fitting of the model parameters was performed in R where binomial proportion 95% confidence intervals (based on the Wilson method) around the model predictions were calculated using the HMisc package (Harrel, 2016). Spearman's Rank Correlation between the frequency of the sequence variants estimated by the model and that of observed was used to determine the statistical significance of model fit, using the MASS package in R (Venables and Ripley, 2003).

The analysis was first performed considering only the sequence variants shared by the source community and the biofilm community. In this case, all sequence variants that were detected in the biofilm, but not in the source, and vice versa were excluded, as it is routinely done (Morris *et al.*, 2013). This approach, however can exclude a significant amount of taxa (in our case less than 10 sequence variants were shared by source and biofilm at NOB guild level), thus reducing confidence in model output. Therefore, we included a second approach, where all sequences present in biofilms, but not detected in the source community, were assigned the lowest observed average relative abundance in the source community (0.001%); i.e., we extended our source community based on the knowledge of our experimental design, where we assume that all sequences observed in the biofilm must have originated from single source (tap water), even when not detected at the time of sampling.

Statistical analyses

All samples were rarified to the same sequencing depth by randomly subsampling 9400 sequences per sample for 10 times for the total community dataset and 5170 sequences per sample in the NOB guild dataset. The α and β diversity calculations were implemented on 10 subsamples of each sample. α-diversity indexes were calculated using algorithms implemented in QIIME, Vegan and Phyloseq packages in R (Oksanen *et al.*, 2011; McMurdie

and Holmes, 2013). The significant differences between the α-diversity indices were assessed with a parametric Student's t-test implemented in QIIME.

A meta community was created for the β diversity significance test by combining related OTU libraries of all treatments and added into the OTU-tables of original samples for further subsampling process. Mean and standard deviation of β diversities, obtained from the distance matrices created with Bray-Curtis algorithm (Bray and Curtis, 1957), were calculated and applied to estimate corrected β diversities as described in Gülay and Smets (2015). All analyses were implemented using R environment with custom codes available at "https://github.com/ardagulay/Beta-Diversity-Significance-Method".

The within-group and between-group dissimilarities were calculated using the metaMDS command in R package Vegan (Oksanen *et al.*, 2011) based on Bray-Curtis distances and plotted with ggplot2 (Wickham, 2009).

Acknowledgments

The authors would like to thank Dr. Olivier Thas and Stijn Hawinkel (Ghent University) for their advice on statistical analysis, Marta Vignola (Newcastle University) for her support with the neutral community assembly model and Yunjie Ma (Technical University of Denmark) for her help during reactor operation.

This research was funded by MERMAID (An Initial Training Network funded by the People Programme - Marie Skłodowska-Curie Actions- of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n°607492).

The authors declare that they have no conflict of interest.

References

Alawi, M., Off, S., Kaya, M., and Spieck, E. (2009) Temperature influences the population structure of nitrite-oxidizing bacteria in activated sludge. *Environ. Microbiol. Rep.* **1**:

184–190.

Bray, R.J. and Curtis, J.T. (1957) An ordination of the upland forest communities of southern Winsconin. *Ecol. Monogr.* **27**: 325–349.

Burns, A.R., Zac Stephens, W., Stagaman, K., Wong, S., Rawls, J.F., Guillemin, K., and Bohannan, B.J. (2015) Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *Isme J* **10**: 1–10.

Daims, H., Lebedeva, E.V., Pjevac, P., Han, P., Herbold, C., Albertsen, M., et al. (2015) Complete nitrification by Nitrospira bacteria. *Nature* **528**: 504–509.

Daims, H., Lücker, S., and Wagner, M. (2016) A New Perspective on Microbes Formerly Known as Nitrite-Oxidizing Bacteria. *Trends Microbiol.* **24**: 699–712.

Evans, S., Martiny, J.B.H., and Allison, S.D. (2017) Effects of dispersal and selection on stochastic assembly in microbial communities. *ISME J.* **11**: 176–185.

Gülay, A. and Smets, B.F. (2015) An improved method to set significance thresholds for β diversity testing in microbial community comparisons. *Environ. Microbiol.* **17**: 3154–3167.

Harrel, F.E. (2016) Hmisc: Harrell Micellaneous. R package version 3.17-3 https://cran.r-project.org/web/packages/Hmisc/index.html.

Koch, H., Galushko, A., Albertsen, M., Schintlmeister, A., Spieck, E., Richter, A., et al. (2014) Growth of nitrite-oxidizing bacteria by aerobic hydrogen oxidation. *Science (80-.).* **345**: 1052–1054.

Koch, H., Lücker, S., Albertsen, M., Kitzinger, K., Herbold, C., Spieck, E., et al. (2015) Expanded metabolic versatility of ubiquitous nitrite-oxidizing bacteria from the genus Nitrospira. *Proc. Natl. Acad. Sci. U. S. A.* **112**: 11371–6.

Langenheder, S. and Szekely, A.J. (2011) Species sorting and neutral processes are both important during the initial assembly of bacterial communities. *Isme J.* **5**: 1086–1094.

Lee, J.E., Buckley, H.L., Etienne, R.S., and Lear, G. (2013) Both species sorting and neutral processes drive assembly of bacterial communities in aquatic microcosms. *FEMS Microbiol. Ecol.* **86**: 288–302.

Lindström, E.S. and Langenheder, S. (2012) Local and regional factors influencing bacterial community assembly. *Environ. Microbiol. Rep.* **4**: 1–9.

Maixner, F., Noguera, D.R., Anneser, B., Stoecker, K., Wegl, G., Wagner, M., and Daims, H. (2006) Nitrite concentration influences the population structure of Nitrospira-like bacteria. *Environ. Microbiol.* **8**: 1487–1495.

McMurdie, P.J. and Holmes, S. (2013) Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**: e61217.

Moroenyane, I., Chimphango, S.B.M., Wang, J., Kim, H.K., and Adams, J.M. (2016) Deterministic assembly processes govern bacterial community structure in the Fynbos, South Africa. *Microb. Ecol.* **72**: 313–323.

Morris, A., Beck, J.M., Schloss, P.D., Campbell, T.B., Crothers, K., Curtis, J.L., et al. (2013) Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am. J. Respir. Crit. Care Med.* **187**: 1067–1075.

Nowka, B., Daims, H., and Spieck, E. (2015) Comparison of oxidation kinetics of nitrite-oxidizing bacteria: Nitrite availability as a key factor in niche differentiation. *Appl. Environ. Microbiol.* **81**: 745–753.

Ofiteru, I.D., Lunn, M., Curtis, T.P., Wells, G.F., Criddle, C.S., Francis, C.A., and Sloan, W.T. (2010) Combined niche and neutral effects in a microbial wastewater treatment community. *Proc. Natl. Acad. Sci.* **107**: 15345–15350.

Oksanen, J., Blanchet, G., Kindt, R., Legendre, P., O'Hara, R.B., Simpson, G., et al. (2011) Vegan: Community ecology package, 2011. *R Packag. Version* 1–17.

Palomo, A., Jane Fowler, S., Gülay, A., Rasmussen, S., Sicheritz-Ponten, T., and Smets, B.F. (2016) Metagenomic analysis of rapid gravity sand filter microbial communities suggests novel physiology of Nitrospira spp. *ISME J.* **10**: 1–13.

Pinto, A.J., Marcus, D.N., Ijaz, Z., Bautista-de los Santos, Q.M., Dick, G.J., and Raskin, L. (2015) Metagenomic Evidence for the Presence of Comammox Nitrospira-Like Bacteria in a Drinking Water System. *mSphere* **1**: e00054-15.

Roguet, A., Laigle, G.S., Therial, C., Bressy, A., Soulignac, F., Catherine, A., et al. (2015) Neutral community model explains the bacterial community assembly in freshwater lakes. *FEMS Microbiol. Ecol.* **91**:.

De Roy, K., Marzorati, M., Negroni, A., Thas, O., Balloi, A., Fava, F., et al. (2013) Environmental conditions and community evenness determine the outcome of biological invasion. *Nat. Commun.* **4**: 1383.

Sloan, W.T., Lunn, M., Woodcock, S., Head, I.M., Nee, S., and Curtis, T.P. (2006) Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ. Microbiol.* **8**: 732–740.

Sorokin, D.Y., Lücker, S., Vejmelkova, D., Kostrikina, N.A., Kleerebezem, R., Rijpstra, W.I.C., et al. (2012) Nitrification expanded: discovery, physiology and genomics of a nitrite-

oxidizing bacterium from the phylum Chloroflexi. *ISME J.* **6**: 2245–56.

Stegen, J.C., Lin, X., Fredrickson, J.K., Chen, X., Kennedy, D.W., Murray, C.J., et al. (2013) Quantifying community assembly processes and identifying features that impose them. *ISME J.* **7**: 2069–79.

Stoddard, S.F., Smith, B.J., Hein, R., Roller, B.R.K., and Schmidt, T.M. (2015) rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.* **43**: D593–D598.

Terada, A., Lackner, S., Kristensen, K., and Smets, B.F. (2010) Inoculum effects on community composition and nitritation performance of autotrophic nitrifying biofilm reactors with counter-diffusion geometry. *Environ. Microbiol.* **12**: 2858–2872.

Venables, W.N. and Ripley, B.D. (2003) Modern Applied Statistics With S Fourth. Springer, New York.

Wang, J., Shen, J., Wu, Y., Tu, C., Soininen, J., Stegen, J.C., et al. (2013) Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus stochastic processes. *ISME J.* **7**: 1310–1321.

Ward, B.B., Arp, D.J., and Klotz, M.G. eds. (2011) Nitrification ASM Press, Washington, DC.

Wickham, H. (2009) ggplot2: Elegant Graphics for Data Analysis Springer-Verlag New York.

Yu, Y., Lee, C., Kim, J., and Hwang, S. (2005) Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnol. Bioeng.* **89**: 670–679.

Supporting information

Supplemental Information is available in the online version of this article at the publisher's web-site:

Supplemental methods

Figure S1: Microbial activity during community assembly measured by nitrite oxidation

Figure S2: Corrected Bray-Curtis distances at a sampling depth of 9400 individuals. The β significance threshold value was estimated with the method in Gülay & Smets (2015). Box plots represent the mean β diversities between different treatments

Figure S3: Phylogenetic analysis showing the affiliation of the most abundant sequences classified as Nitrospira and Nitrotoga

Figure S4: Fit of the neutral model to observed biofilm community assembly based on total community composition

Figure S5: Fit of the neutral model to observed biofilm community assembly based on NOB guild composition

Table S1: Total bacteria cell numbers for individual reactors quantified by qPCR

Table S2: Total bacteria cell numbers for tap water samples at 6 different time points during the experiment, quantified by qPCR

Table S3: Significance of observed differences in α-diversity, assessed using parametric T-test; given a significance value of 0.05

Table S4: *Comparison of NCM fit between all treatments by including or excluding non-detected sequence variants based on detection limits and testing the hypothesis (H_0) that there is no correlation between model prediction and observed data*

Table S5: All exact sequence variants included in the NOB guild; representative sequence and the classification according to SILVA123 database (see excel file).

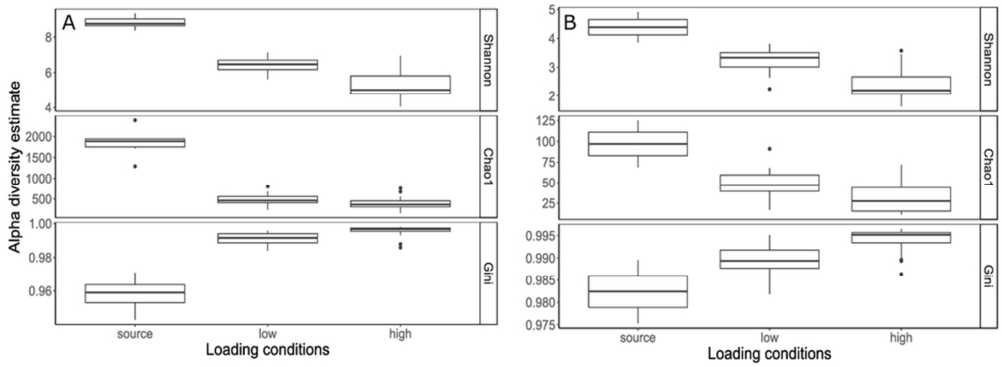
List of tables and figures:

Figure 1: Diversity, richness and evenness estimates of the total community (A) and NOB guild (B) in source and biofilms, determined after normalizing the 16S rRNA gene sequence libraries to 9400 and 5170 sequences, respectively

Figure 2: Community composition of all replicate communities (source and high- and low-nitrite loading biofilms) displayed as relative abundance of the 16S rRNA gene sequences classified at the order level against the Silva123 prokaryotic reference database

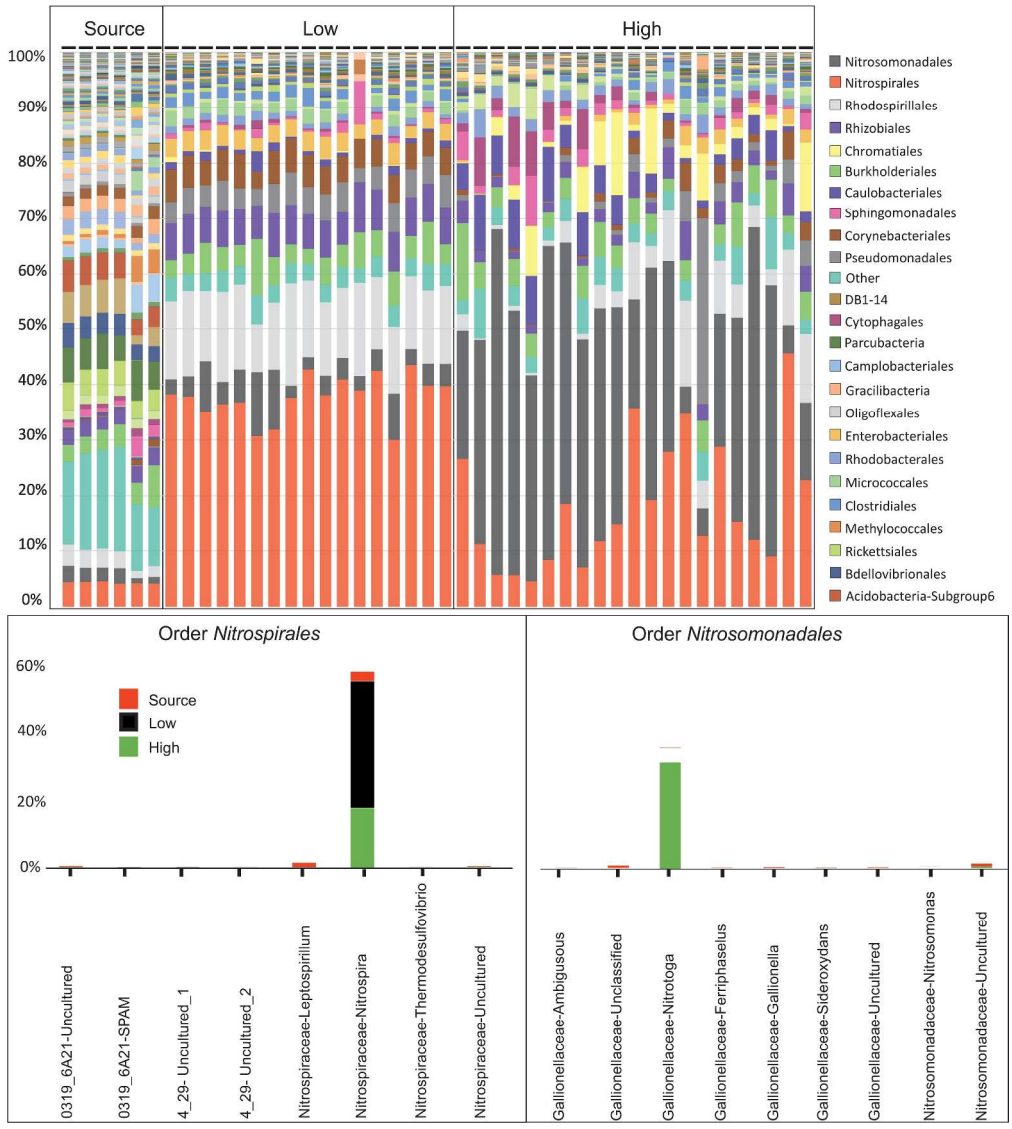
Figure 3: Relative abundance of *Nitrospira* and *Nitrotoga* in biofilm communities assembled under two different nitrite loadings, compared to the source community, based on 16S rRNA amplicon sequences classified at the genus level and *Nitrospira* relative abundance based on *Nitrospira* 16S rRNA specific qPCR.

Figure 4: Nonmetric multidimensional scaling ordinations of replicate communities based on Bray-Curtis distances showing total community (A) and NOB guild (B) biofilm communities at low (green) and high (red) nitrite loadings compared to source community (blue, tap water community)



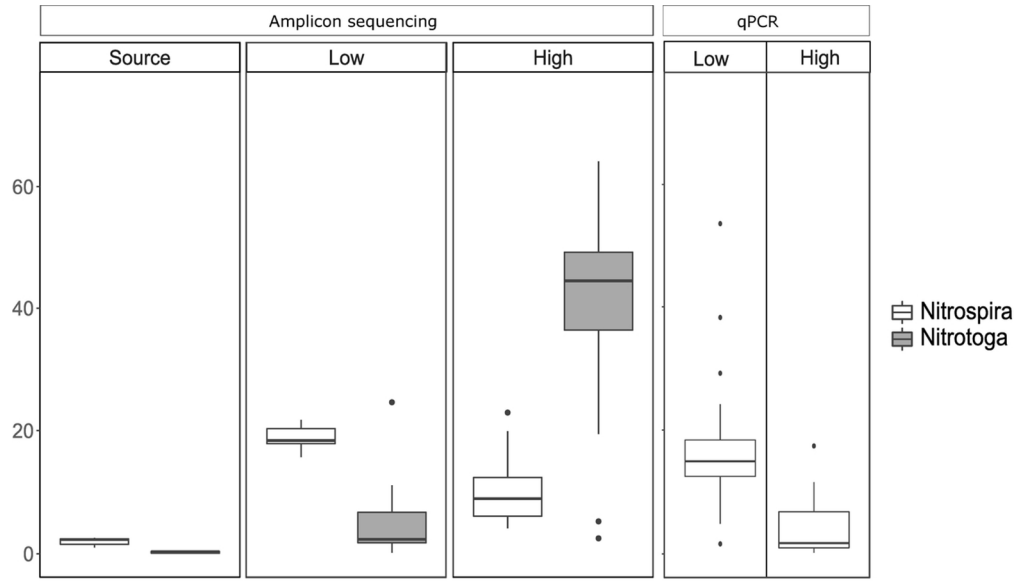
Diversity, richness and evenness estimates of the total community (A) and NOB guild (B) in source and biofilms, determined after normalizing the 16S rRNA gene sequence libraries to 9400 and 5170 sequences, respectively

82x30mm (300 x 300 DPI)



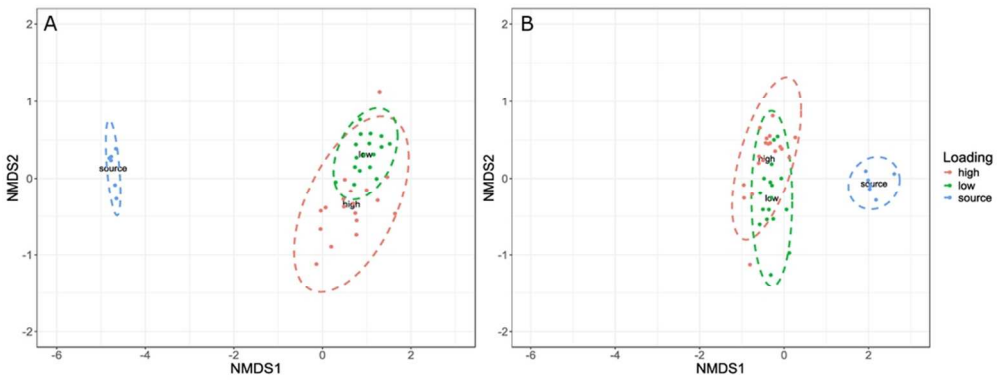
Community composition of all replicate communities (source and high- and low- nitrite loading biofilms) displayed as relative abundance of the 16S rRNA gene sequence variants classified at the order level against the Silva123 prokaryotic reference database

327x365mm (300 x 300 DPI)



Relative abundance of Nitrospira and Nitrotoga in biofilm communities assembled under two different nitrite loadings, compared to the source community, based on 16S rRNA amplicon sequences classified at the genus level and Nitrospira relative abundance based on Nitrospira 16S rRNA specific qPCR

114x65mm (300 x 300 DPI)



Nonmetric multidimensional scaling ordinations of replicate communities based on Bray-Curtis distances showing total community (A) and NOB guild (B) biofilm communities at low (green) and high (red) nitrite loadings compared to source community (blue, tap water community)

86x32mm (300 x 300 DPI)